Supplementary Material for PanoHair: Detailed Hair Strand Synthesis on Volumetric Heads

Shashikant Verma
shashikant.verma@iitgn.ac.in
Shanmuganathan Raman
shanmuga@iitgn.ac.in

CVIG Lab
Indian Institute of Technology
Gandhinagar
India

1 PanoHead

Why PanoHead Over Existing Approaches? Various works have explored implicit volumetric representations of the human head. GRAM [3] utilizes structured manifolds, while [2] employs Tri-Plane representations. PanoHead [1] further advances this by leveraging Tri-Grid feature representations. As illustrated in Figure 1, methods trained solely on frontal views struggle to accurately capture the complete 360° geometry of the human head, posing a challenge for precise hair volume estimation. Figures 1(a) and 1(b) depict the geometric reconstructions obtained by [3] and [2], respectively. PanoHead [1] overcomes this limitation by utilizing tri-grid feature representation $\mathcal{T}_f$, enabling the reconstruction of the full 360° head geometry, as shown in Figure 1(c). However, PanoHead is trained to predict the color and density distribution within a volumetric representation by design, enabling view-consistent neural radiance field rendering. This radiance-based implicit representation, modeled as a volumetric density field, is optimized for rendering but suboptimal for surface extraction. Therefore, we propose to distill knowledge from PanoHead by training $\Psi(.)$ to directly predict signed distances. By leveraging the signed distance function (SDF), the surface can be naturally extracted as the zero-level set or iso-surface, providing a more precise and well-defined geometry. Moreover, we extend $\Psi(.)$ to also predict semantic masks and 3D orientations for every point on the iso-surface of the predicted SDF, enabling a more structured understanding of hair geometry. This comprehensive approach ensures that PanoHair captures both the geometry and semantics necessary for high-fidelity hair modeling.

The ability of tri-grid representations to capture full-head geometry, as demonstrated by PanoHead [1], forms the foundation of PanoHair. Figure 1 qualitatively compares head geometries obtained from various 3D-aware neural representations. Generative Manifolds [3] and Tri-planes [2] primarily reconstruct frontal geometries (Figure 1(a), (b)). In contrast, we leverage knowledge distillation from PanoHead to learn full-head SDFs (Figure 1(d)), whereas PanoHead represents faces as density distributions (Figure 1(c)).

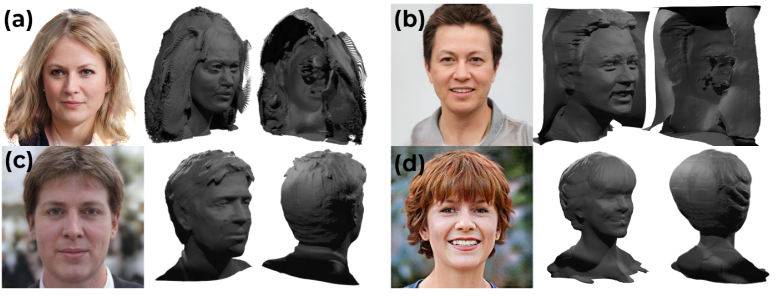

Figure 1: The geometric shapes extracted from implicit representations learned using existing methods: (a) GRAM [3], (b) Tri-Planes [2], (c) PanoHead [1], and (d) Ours.

2 Loss Functions

In the main paper, we define our loss function $\mathcal{L}$ as presented in Equation 1

$$\mathcal{L} = \underbrace{\lambda_1 \cdot \delta(t) \cdot \|\mathcal{D} - \hat{\mathcal{D}}\|_1}_{\textbf{Density Loss}} + \underbrace{\lambda_2 \cdot \|I^+ - I\|^2}_{\text{Image Reconstruction Loss}} + \underbrace{\lambda_3 \cdot \text{BCE}(I_M, \hat{M}_{2D})}_{\textbf{Semantic Loss}} + \mathcal{L}_{\text{orient}} \tag{1}$$

Here, λ represents the weights assigned to each loss term, and $\mathcal{L}_{orient}$ denotes the orientation loss, which we now describe in detail.

2.1 Tangential Loss.

Since the SDF $\mathcal{S}$ encodes signed distances, its gradient at the surface points x_S points in the direction of the outward surface normal. We impose tangential loss for orientations o_S predicted at surface points as defined in equation 2.

$$\mathcal{L}_{\text{tan}} = \frac{1}{N} \sum_{i=1}^{N} |o_S^i \cdot n_S^i| \quad \text{where} \quad n_S = \frac{\nabla \mathcal{S}}{||\nabla \mathcal{S}||} \tag{2}$$

2.2 Multi-view Orientation Projection Loss.

We generate k multi-view outputs from PanoHead using cameras ${}^iC_{\text{cam}}$ and render the corresponding super-resolved images I^+ along with their Gabor orientation maps I_O. For clarity, we adopt the notation ${}^iX$, where the superscript i on the top left of a variable X indicates its association with the i-th view. To ensure consistency across views, we synthesize these k views so that each image overlaps by more than 70% with the first view. For surface points ${}^0x_S$ identified in the first view, we project them into all k camera views, including the first view itself, and infer their ground-truth Gabor orientations from ${}^iI_O$. Gabor orientation maps represent local image structures by capturing dominant edge directions. However, they exhibit directional ambiguity because an orientation vector θ and its opposite direction $\theta + \pi$ produce the same response. This means that for a predicted orientation ${}^0\hat{\mathcal{O}}_{2D}$ its true counterpart I_O can be any of the $I_O, I_O + \pi$, and $I_O - \pi$. Hence, for projected predicted orientations

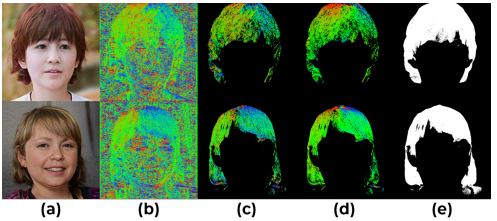

Figure 2: For the image (a) generated from a latent code, (b) shows estimated Gabor orientations, while (c) highlights high-confidence (> 0.2) orientations in the hair region. Predicted orientations for hair region points are shown in (d), and screen-projected locations of pixels belonging to hair semantics in (e).

Loss	L_{tan}	L_{proj}	L_{tan}	L_{proj} (rad)
Multi-view	✓	✓	0.16	0.31
Multi-view	✗	✓	0.20	0.34
Single-view	✓	✓	0.18	0.42
Single-view	✗	✓	0.26	0.48

Table 1: Impact of training PanoHair with and without multiple-views in consideration

${}^0\hat{\mathcal{O}}_{2D}$ in the first view, we impose projection loss as in Equation 3. This enables a multi-view consistent optimization of the orientation vector.

$$\mathcal{L}_{\text{proj}} = \sum_{i=0}^{k-1} \min\Big(\left|{}^0\hat{\mathcal{O}}_{2D} - {}^iI_O\right|, \left|{}^0\hat{\mathcal{O}}_{2D} - {}^iI_O - \pi\right|, \left|{}^0\hat{\mathcal{O}}_{2D} - {}^iI_O + \pi\right| \Big) \tag{3}$$

Finally, the total orientation loss is defined as in equation 4.

$$\mathcal{L}_{\text{orient}} = \lambda_4 \mathcal{L}_{\text{tan}} + \lambda_5 \mathcal{L}_{\text{proj}} \tag{4}$$

We train PanoHair for 1 million iterations, sampling latent codes z from a normal distribution $p_z(0,1)$, using the loss functions defined in Equation 1. For multi-view projection loss, we randomly sample $k = 3$ multiple overlapping views. The weighting parameters are set as follows: $\lambda_1 = 1e-4$ (decaying to 0 after 20K iterations), $\lambda_2 = 1$, $\lambda_3 = 10$, and $\lambda_4 = \lambda_5 = 10$. Training is conducted on an NVIDIA RTX-4090 GPU and takes approximately three days. Figure 2 shows predicted orientations compared with ground-truth Gabor orientations along with $\hat{M}_{2D}$. In Table 1, we present an ablation study analyzing the impact of training PanoHair with a single-view orientation projection loss versus a multi-view projection loss. For each setting, we also evaluate the effect of omitting the tangential loss. The results show that the model trained with both the tangential and multi-view projection losses achieves the best performance, yielding the lowest angular error when projecting predicted orientations onto the ground truth. Moreover, the predicted orientations remain more tangential to the surface of the hair mesh, which is desirable. The evaluation is conducted by randomly sampling 100 latent codes.

References

[1] Sizhe An, Hongyi Xu, Yichun Shi, Guoxian Song, Umit Y. Ogras, and Linjie Luo. Panohead: Geometry-aware 3d full-head synthesis in 360deg. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20950–20959, June 2023.

[2] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16123–16133, 2022.

[3] Yu Deng, Jiaolong Yang, Jianfeng Xiang, and Xin Tong. Gram: Generative radiance manifolds for 3d-aware image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10673–10683, 2022.